DOCUMENT RESUME

ED 255 012 EC 172 071

AUTHOR Strum, Irene
TITLE Alternative Testing Techniques: Impact and Meaning.
PUB DATE 3 Apr 85
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Disabilities; Elementary Education; *Individualized Education Programs; *Reading Tests; Testing Problems; *Urban Education
IDENTIFIERS *Adapted Tests (for Disabled); New York (New York)

ABSTRACT

More than 50,000 New York City special education students in self-contained classes and resource rooms in grades 2-9 were tested in reading on the California Achievement Test, Form C. Students who had testing modifications recommended on their individualized education programs were tested with the appropriate modification. Among the moderately impaired students in self-contained classes, the positive effect of testing modifications seemed greatest among lowest scoring students. Only three percent more students tested with modifications scored at or above grade level than those tested without modifications, but six percent more who were not tested with modifications scored two or more years below grade level than students tested with modifications. However, it was found that, outside of this study, many students for whom modifications were authorized were tested without them. Frequently principals did not provide a testing environment conducive to alternative testing techniques. Lack of sufficient staff was also a problem. It was concluded that although modifications of tests for handicapped students has generally been accepted, significant problems remain. (CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED255012

ALTERNATIVE TESTING TECHNIQUES: IMPACT AND MEANING

Irene Strum
New York City Public Schools
Division of Special Education

Paper Presented at the 1985
AERA Annual Meeting
Chicago, Illinois,
April 3, 1985

EC172071

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Irene Strum
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A history of assessment would emphasize the selective and discriminatory nature of testing. Kamen (1974) noted the importance of distinguishing between those who can, who have the ability, or who perform the task from the cannots and have nots, regardless of environmental factors. Tests which were culturally biased or which included items outside the experience of the subject were widely used, even in the post World War II period. As late as the 1960's, general ability tests, known as they are today as scholastic aptitude tests, were defended as predictors of academic achievement even if culturally biased. The explanation emphasized the need for conformity to the mainstream culture if academic success was to be realized. As a result, tests became walls and locked gates for some, keeping out individuals rather than broadening opportunity.

In the sixties, research showed how normative samples were unrepresentative of populations for whom the tests were used. Urban centers were traditionally under- or unrepresented. Moreover, bias was found not only in standardization programs, but in item selection. Green (1972) found that item tryouts were frequently implemented among homogeneous groups of subjects living in middle income suburban areas. Resulting test items were biased against minorities.

In the 1970's and 1980's major test publishers have corrected their procedures to include many more students who are representative of the population as reported in the United States census. Moreover,

items are frequently tried out with minority populations. But what about the handicapped? Special educators find that the newest tests include some mildly handicapped students in the normative group. These students are generally resource room students who are mainstreamed for most of the school day. Sometimes validity studies are included in the test manual to show results of the test for small groups of more severely handicapped students who have been placed in self-contained classes.

Educators and test developers recognize the need to ensure fairness in testing. For example, test questions are reviewed and empirically tested to remove items that test concepts or language which may be biased against or insensitive to minorities. Test administration must be regulated to assure similarity of test conditions. Test users are informed on how to use test scores. Studies are periodically undertaken to compare the performance of different groups. But for some goups of test takers, such as the handicapped, "individual needs vary so dramatically that testing conditions cannot even approach uniformity" (ETS, 1983, p. 22).

PL94-142 reinforced the need to include the handicapped in test administrations. However, to facilitate widespread participation in testing programs, it was necessary to modify test conditions. Although the tests were originally developed for general education, they are needed to measure the skills and abilities of students in special education. Practictioners have focused on how to test

individuals whose handicapping condition interferes with taking the test and how to measure their abilities and skills with a minimum influence of their handicapping condition.

The need for ensuring civil rights for the handicapped has encouraged most states to mandate alternative testing techniques or test modifications for students with handicapping conditions (Gillespie, 1982; Schenck, 1981; University of the State of New York, 1982). In a survey conducted by Strum and Ribner (1984), it was reported that almost all states and urban centers mandated test modifications for students with handicapping conditions. In addition, test developers, such as Educational Testing Service (ETS), have authorized the use of modifications for handicapped individuals in major national testing programs. These adaptations have been accepted although Rosewater (1979) and Sherman and Robinson (1982) acknowledged that their effect on reliability and validity is generally unknown.

In fact, Willingham has stated that "it is difficult to say whether (modifications)...are adequate or could compensate for impairments that tend to lower test performance" (ETS, 1983, p.22). As a result, ETS is undertaking a number of studies, the goal of which is to develop scales which will correct for biases operating against the handicapped. Florida and North Carolina are also undertaking studies to determine the effect of modifications on test scores. Empirical research conducted in Florida (Beattie, 1983; Grise, 1982) reported that learning disabled students obtained somewhat higher scores when tested

with modifications than when tested without them. However, the modifícation consisted primarily of a different test format.

Generally, all modifications must be indicated on the student's Individual Education Plan (IEP). In New York, they must be reflected consistently in the IEP and should be an "integral and ongoing part of the...IEP" (University of the State of New York, 1982 p.2); Moreover, the modifications" should be necessary and consistently utilized by the pupils for any authorized testing situation...throughout the course of his or her instructional program" (University of the State of New York, 1982, p.2).

New York City implements the modifications mandated by New York State. These are:

1. extended or waived time limits
2. special location
3. reading questions aloud
4. answering in any manner
5. using a calculator or abacus
6. Braille or large type examination copy
7. questions omitted and credit pro-rated

Extended and/or waived time limits and testing in a special location change test procedures, but not the skills being tested. The other modifications change testing procedures so that the skills tested are different from those originally intended by the test makers. A reading comprehension test may become a listening comprehension test. Writing or written expression may become oral expression. Mathematics competency may become problem solving ability. However, the abilities that are tested in a test may not be as appropriate for the handicapped as a related ability which requires similar underlying mental processes, i.e., comprehension, expression.

Extended and/or waived time limits include extra working time as well as short breaks after short periods of working. In New York City, extended time is limited by administrative regulations requiring prompt submission of answer documents. Test security also demands short breaks during test administration with no participation of students in activities with other students who have completed the test. Although some states and cities require indication of the actual number of minutes for time extension on the IEP, others permit very general statements of need.

The Joint Technical Standards for Educational and Psychological Testing (APA, 1984) will probably indicate that "there are few data available to support any conclusions about the effects of modifications in time, number of settings, or number of recesses on the test results. Further, little is known about how much time people with various handicapping

conditions actually need because records of time actually used are rare, and empirical studies to set time limits are even less common" (p.6-1, 6-2). Although this modification is frequently criticized on the basis of violating standardization procedures, it should be noted that most achievement tests are described by test publishers as power tests in that 90% of the students taking the test complete it. The special education student is generally among the 10% who are unable to complete the examination during prescribed time limits. For these students, the test is a speed test, not a power test. Increasing time limits may, therefore, improve test reliability for the individual student with a handicapping condition, rather than maintaining the time as indicated in the manual.

Administering an examination in a special location generally refers to test administration in a hospital or home. However, for some students in special education, it may mean decreasing class size for test administration. Sometimes students are tested one-to-one in a small office. It definitely does not mean enlarging the group and taking it to the auditorium or lunchroom for test taking.

Reading questions aloud on a math test has the effect of removing reading ability from the math score. This procedure increases the validity of math test scores for students who have reading

disabilities. Reading questions aloud on a reading test reduces the frustrations of students who have short attention spans. When the passage as well as the questions are read aloud, the test measures a skill that is different from reading. Thus, on the New York State competency tests, PEP, PCT, and RCT, in reading, students who have the test read aloud are being assessed for listening comprehension rather than reading comprehension. However, the underlying processes for both reading and listening comprehension have been deemed identical. On the New York City reading test a typical global instrument measuring vocabulary, comprehension, and total reading, only the questions are read aloud, not the passage.

Answering questions in any manner is appropriate for students who are unable to use a separate answer document. Students with disabilities affecting eye-hand coordination and visual preception frequently respond orally to an amanuensis who fills in the appropriate space on the answer document. In the State writing competency tests, RCT, PEP, or PCT, students who dictate their responses to an amanuensis are being tested in oral expression rather than written expression.

Use of a calculator or abacus are the only modifications in math tests which change the skill being tested. Students who use these devices do not exhibit knowledge of math facts and operations. However, they do need reasoning ability and problem solving skills.

Braille or large type examination copy permits many students to participate in tests who would otherwise be prevented from participation by the nature of their disability. Generally, questions are omitted and credit pro-rated for students for whom Braille or large type copy are not available. In addition, some questions include presentations of information, such as pictures or graphs, which cannot be Brailled or enlarged. In these instances, test items must be omitted. This modification is particulary difficult to implement in large scale testing programs because scanning/scoring equipment cannot be programmed easily for the few students who require it.

It is interesting to consider that the use of a tactile approach to reading by blind students taking a reading examination in Braille is rarely if ever criticized as measuring a trait other than reading comprehension. In fact, modifications have been accepted fairly extensively for students with physical disabilities. The acceptance of modifications for students whose handicapping conditions are less obvious is generally questioned because of the lack of understanding of the need of these students. Modifications are not based on lack of ability or achievement. Students placed in special education solely on the basis of low intellectual ability are not eligible for most modifications. A student who is eligible for a modification has a special need for it in order to obtain a "true" measure of the ability or achievement being measured with minimal contribution of the disability. The appropriate use of modifications should preserve the nondiscriminatory character of the testing program.

Method

In spring 1983 special education and general education students were tested in reading in New York City. More than 50,000 special education students in self-contained classes and resource rooms in grades two through nine participated. The reading test selected was the California Achievement Test (CAT), Form C. Students who had testing modifications on their IEP were tested with the appropriate modification(s).

Students in grades 2 and 3 were given machine scorable test booklets. All other students used separate answer documents. All materials were scanned by New York City public school equipment and scores were reported on tape and hard copy by the Office of Student Information Services.

Findings

Among the moderately impaired students in self-contained classrooms, the positive effect of modifications seemed greatest among lowest scoring students; only three percent more students tested with modifications scored at or above grade level than thosetested without modifications, but six percent more who were not tested with modifications scored two or more years below grade level than students with modifications.

TABLE I

FREQUENCY DISTRIBUTION OF READING ACHIEVEMENT AMONG STUDENTS TESTED IN SELF-CONTAINED CLASSROOMS GRADES 2-9 IN RELATION TO TEST ADMINISTRATION PROCEDURES 1983

	TOTAL		TESTED WITH MODIFICATIONS		TESTED WITHOUT MODIFICATIONS	
	NO	%	NO.	%	NO.	%
At or Above Grade Level	4432	13	3248	12	1184	15
Less Than 1 Year Below Grade Level	5802	17	4615	18	1187	15
One to Less Than 2 Years Below Grade Level	10286	30	8231	31	2055	26
Two or More Years Below Grade Level	13740	40	10147	39	3593	45
TOTAL	34260	100	26241	100	8019	100**

** Total does not add to 100% because of rounding.

Among resource room students it was found that little if any differences existed between students who were tested with modifications and without.

TABLE 2

FREQUENCY DISTRIBUTION OF READING ACHIEVEMENT AMONG STUDENTS TESTED IN RESOURCE ROOMS GRADES 2-9 IN RELATION TO TEST ADMINISTRATION PROCEDURES 1983

	TOTAL		TESTED WITH MODIFICATIONS		TESTED WITHOUT MODIFICATIONS	
	NO.	%	NO.	%	NO.	%
At on Above Grade Level	6639	33	4194	34	2445	31
Less Than 1 Year Below Grade Level	6453	32	3863	32	2590	32
One to to Less Than 2 Years Below Grade Level	4963	25	2690	24	2073	26
Two or More Years Below Grade Level	2131	11	1229	11	902	12
TOTAL	20186	100**	12176	100**	8010	100**

**Total does not add to 100% because of rounding.

The records of resource room students in grades two through nine were checked to determine which students were eligible for modifications. The 1983 reading scores of students who were eligible for modifications and tested with them were compared with those eligible but for whom modifications were not implemented.

It was found that mean differences between the groups at various grade levels were slight. In some grades, the modified group scored higher; in others, those without modifications scored somewhat higher.

TABLE 3

MEAN READING ACHIEVEMENT

GRADE EQUIVALENTS BY GRADE AND COMMUNITY SCHOOL DISTRICT AMONG RESOURCE ROOM STUDENTS IN RELATION TO TEST PROCEDURE

GRADE	2		3		4		5		6		7		8		9	
DISTRICT	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.
1	1.9	2.1	2.7	2.5	4.1	3.8	5.0	4.9	6.3	5.9	6.8	6.8	8.6	8.6	7.8	7.7
2	2.5	2.1	3.2	3.1	4.2	4.9	6.1	5.4	6.2	6.6	6.7	6.5	9.2	8.1	9.2	10.2
3	1.9	2.2	2.6	2.9	4.0	4.3	5.1	5.0	5.7	6.2	7.2	6.6	8.9	7.6	8.7	8.1
4	2.7	2.7	3.3	3.2	4.3	3.8	5.0	4.8	6.2	5.7	6.5	5.1	7.4	7.0	10.5	8.5
5	2.5	2.7	3.1	2.3	4.2	4.3	5.0	5.2	4.8	5.7	7.7	7.1	8.7	7.5	7.8	8.0
6	2.0	1.6	3.1	2.4	3.5	3.9	5.0	4.6	5.0	4.7	6.7	6.5	7.4	6.8	8.7	----
7	2.5	1.7	2.4	2.7	3.9	3.7	5.0	4.7	5.2	5.6	6.4	6.3	7.8	6.7	---	7.5
8	1.9	2.4	3.0	3.2	4.1	4.1	5.4	5.1	5.5	6.2	6.9	7.2	8.5	8.6	9.1	8.3
9	1.6	2.1	3.0	3.0	3.6	4.2	4.8	5.4	5.8	6.2	7.2	5.7	8.0	7.7	---	----
10	2.4	2.2	3.3	2.8	3.9	3.8	5.8	4.8	6.0	5.0	6.6	6.8	8.1	7.8	8.1	9.0
11	2.6	2.5	3.3	3.1	4.9	4.3	5.5	5.3	6.7	6.5	7.5	8.0	8.3	8.5	9.1	10.2
12	2.0	2.3	2.8	3.5	3.6	3.8	5.2	5.3	4.9	6.4	8.3	6.9	7.8	8.2	---	----
13	2.7	2.6	3.6	3.1	4.4	4.3	5.6	4.7	6.4	5.9	7.5	6.8	7.7	7.3	---	----
14	1.8	2.2	2.5	2.8	3.8	4.2	4.3	5.5	6.0	5.2	6.6	6.4	9.0	7.4	---	7.6
15	2.5	2.3	2.8	3.0	4.3	4.2	5.3	6.0	6.0	6.0	6.8	7.2	8.4	8.2	9.0	9.0

BEST COPY AVAILABLE

GRADE	2		3		4		5		6		7		8		9	
DISTRICT	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.	MOD.	NO. MOD.
16	2.3	1.1	3.0	2.8	3.8	3.9	4.9	4.5	6.7	5.5	6.4	6.1	7.4	7.8	6.7	6.9
17	2.8	2.6	2.9	3.1	3.7	3.8	5.3	5.0	5.3	5.4	6.8	6.3	8.2	7.3	---	---
18	2.1	2.8	3.6	3.0	4.3	4.1	5.4	5.2	6.2	5.9	7.0	7.4	8.6	7.8	9.1	8.9
19	2.2	2.0	2.8	2.8	4.0	4.3	5.0	5.1	5.5	5.8	6.7	6.7	7.2	8.3	---	7.3
20	2.5	2.3	3.4	2.6	4.3	4.2	5.2	5.1	6.3	6.1	7.3	6.9	8.4	8.7	8.9	7.5
21	2.4	2.4	3.0	3.2	4.5	4.1	5.6	5.2	6.6	6.6	7.5	7.2	8.3	9.1	9.1	7.6
22	2.4	2.4	3.4	3.6	4.7	4.5	5.8	5.9	6.4	6.3	7.3	7.7	8.6	9.1	9.3	11.3
23	1.9	2.2	2.4	3.4	3.8	4.0	5.0	4.5	5.7	6.3	6.4	6.9	6.9	8.4	---	---
24	1.7	2.2	2.9	3.2	4.5	4.3	5.8	5.9	6.3	6.0	7.3	7.0	8.3	9.1	9.9	10.5
25	2.4	2.7	3.4	3.5	4.7	5.2	5.8	5.9	7.1	6.6	8.3	7.7	8.7	9.0	9.5	9.2
26	2.3	2.2	3.4	3.6	4.7	5.0	5.9	5.9	6.7	7.2	8.0	7.6	9.2	8.9	9.5	9.7
27	2.1	2.3	2.8	2.9	4.1	3.9	5.1	5.1	5.9	6.0	6.9	8.4	7.9	8.1	9.8	----
28	2.0	2.3	3.2	3.0	4.4	4.4	5.6	5.6	6.2	6.4	7.2	7.2	8.8	8.9	9.1	9.3
29	2.4	2.7	3.2	3.5	4.1	4.6	5.3	5.8	6.5	6.3	7.0	6.8	8.2	8.2	9.1	9.3
30	2.4	2.4	2.9	3.0	4.4	4.3	5.5	5.5	6.0	5.9	7.0	6.4	8.9	9.0	10.9	10.0
31	2.3	2.3	3.1	3.3	4.8	4.5	5.7	5.4	6.1	5.9	7.5	6.7	8.7	8.4	---	----
32	2.1	2.4	2.7	2.8	4.2	4.0	5.1	5.0	5.1	5.2	7.3	6.3	9.0	8.2	---	9.5
$SE_{\bar{x}}$	.25		.33		.44		.54		.60		.77		.67		.72	

NOTE: N's generally varied from 22 to 100 with city totals reaching a minimum of 500.

17

16

To determine whether differences were significant or educationally important, they were compared with the standard error of of measurement (SE_T), which indicates the range of individual's score. These comparsions showed that the differences were smaller than the SE_T in more than half the districts.

TABLE 4

NUMBER OF DISTRICTS WITH DIFFERENCES BETWEEN RESOURCE ROOM STUDENTS TESTED WITH MODIFICATIONS AND THOSE WITHOUT MODIFICATIONS IN RELATION TO GRADE

GRADE	2	3	4	5	6	7	8	9*
NUMBER OF DISTRICTS WHERE: STUDENTS TESTED WITH MODS WERE HIGHER	12	12	15	18	16	21	16	9
STUDENTS TESTED WITHOUT MODS WERE HIGHER	15	18	15	10	14	8	14	9
NUMBER OF DISTRICTS WITH: NO DIFFERENCE	5	2	2	4	2	3	2	1
DIFFERENCES WERE GREATER THAN SE_T	16	7	6	6	7	6	12	9
IN FAVOR OF TESTED WITH MODS	4	6	2	3	3	4	8	5
IN FAVOR OF TESTED WITHOUT MODS	12	1	4	3	4	2	4	4

*Many 9th graders are in high schools and are not tested in this program.

Discussion

The real effect of modifications, as indicated above, is unknown. The large N and the lack of equivalent or randomly assigned groups precluded the use of parametric statistics. Collected data could not be subjected to inferential statistics, but nonetheless yielded some indications of the effects of modifications.

Although teachers were supposed to test students according to the test procedures indicated on their IEPs, it was found that many students for whom modifications were authorized were tested without them. Frequently principals did not provide a testing environment which was conducive to alternative testing techniques. Lack of sufficient staff did not permit the implementation of modifications to students. Some teachers may not have checked IEPs to identify students for whom they were appropriate.

It should be noted that almost all resource room students who were tested with modifications were eligible for extended time or waived time limits. No more than nine percent were eligible for the more extreme "reading questions aloud" and "answering in any manner."These findings support the hypothesis that "modifications are directly related to handicapping conditions in that

their use permits students to participate in programs from which they otherwise would be excluded" (Strum, 1983, p. 22).

Although modification of tests for handicapped students has generally been accepted, significant problems remain. The effects of alternative testing techniques on resulting scores and on their reliability and validity have not been studied empirically. Use of some modifications appears to change the constructs measured by the test. To what extent do these changes affect the measurement of cognitive functioning and the meaning of the resulting scores? Research on the comparability of scores resulting from modified testing and traditional approaches is necessary to ensure fairness to all individuals tested. Until empirical evidence demonstrates the biased and discriminatory nature of modified test procedures for students with handicapping conditions, the current policy of implementing modifications should be strongly supported.

REFERENCES

American Psychological Association (1984). Joint technical standards for educational and psychological testing. Fourth Draft.

Beattie, S. (1983). Effects of test modifications on the minimum competency performance of learning disabled students. Learning Disability Quarterly, 6, 75-77.

Educational Testing Service (1983). Annual report: Princeton, NJ: ETS.

Gillespie, E.B. (1982). Minimum competency testing for special education students: Issues on implementation: Houston: Workshop T4 at CEC Annual Convention.

Green, R. (1972). Racial and ethnic bias in test construction. Final report, USOE contract No. OEC 9-70-0058 (057). Monterey, CA: CTB/McGraw-Hill.

Grise, P. (1982). Assessment of minimum competency in fifth grade learning disabled students: Test modifications make a difference. Journal of Educational Research, 76, 35-40.

Kamen, L. (1974). The science and polities of I.Q. Hillsdale, NJ: Erlbaum

Rosewater, A. (1979). Minimum competency testing program and handicapped students: Perspective on policy and practice. Washington, D.C.: Institute for Educational Leadership, Georgetown University.

Schenck, S.J. (1981). Ramifications of the minimum competency movement for special education. Los Angeles: Paper presented at AERA annual meeting.

Sherman, S.W., & Robinson, N.L. (1982). Ability testing of handicapped people: Dilemma for government, science, and the public. Washington, D.C.: National Academy Press.

Strum, I. (1983). Reading achievement, Grades 2-9 division of special education report on CAT results, April, 1983. New York City: New York City Public Schools.

Strum, I., & Ribner, S. (1984). Survey on testing students with handicapping conditions. New York City: New York City Public Schools.

University of the State of New York (1982). Alternative testing techniques for children with handicapping conditions. Albany, New York: University of the State of New York, the State Education Department, Office for Education of Children with Handicapping Conditions.